DOCUMENT RESUME

ED 081 811 TM 003 172

AUTHOR Lichtman, Marilyn
TITLE The Development and Validation of R/EAL, an Instrument to Assess Functional Literacy.
SPONS AGENCY Catholic Univ. of America, Washington, D.C.; Manpower Administration (DOL), Washington, D.C. Job Corps.
PUB DATE Jun 73
NOTE 38p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Adolescents; Adults; Criterion Referenced Tests; Data Analysis; Disadvantaged Groups; *Functional Illiteracy; Measurement Instruments; *Reading Tests; Research Methodology; Tables (Data); *Test Construction; *Test Reliability; Tests; *Test Validity
IDENTIFIERS *Reading Everyday Activities in Life; REAL

ABSTRACT

Research procedures used in the development and validation of R/EAL (Reading/Everyday Activities in Life), a new test to overcome problems in assessing functional literacy among adolescents and adults, are described. Specific objectives of the study were to: (1) provide information about the design and development of R/EAL, including determination of reading criteria, establishment of task analyses, production and selection of individual test items, development of test format and procedures, and completion of the final version of R/EAL; and (2) provide information about the validation of R/EAL, including procedures and data on reliability and validity for a select sample. Following field testing of the R/EAL, a revised version of the instrument was developed, which contains reading criteria selected from common daily reading materials; includes tasks related to each criterion developed based on the reading functions required to deal with the individual criteria; and presents all directions and questions in aural form via individually operated cassette players. The methodology used for validating R/EAL, including sample selection, testing procedures, and item analysis, is described, and data relating to individual item and total test statistics, factor analysis data, reliability, and validity are reported. The reliability and validity figures on R/EAL tend to support its use as a viable assessment instrument for functional literacy. Item analysis figures show a difficulty level of items for the total group raining from .35 to .97 with a median of .60. (DB)

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SEP 18 73

ED 081811

TM 003 172

THE DEVELOPMENT AND VALIDATION OF R/EAL, AN INSTRUMENT TO ASSESS FUNCTIONAL LITERACY

By

Marilyn Lichtman, Ed.D.
The Catholic University of America

Washington, D. C. 20017
June 1973

FILMED FROM BEST AVAILABLE COPY

THE DEVELOPMENT AND VALIDATION OF R/EAL, AN INSTRUMENT TO ASSESS FUNCTIONAL LITERACY*

Marilyn Lichtman, Ed.D.

The Catholic University of America

Recently, interest in literacy in the United States has undergone serious changes in emphasis and direction. Increased emphasis on the extent of illiteracy was spawned by Allen's Right to Read speech of the sixties resulting in an ever-increasing public awareness that some 18½ million adult Americans remain basically incapable of performing simple tasks involving minimum reading skills. Federal, state and local efforts have been mounted to deal with the problem. Right to Read Councils are mushrooming across the country in both schools and communities. A renewed interest in reading and literacy has been taken in Congress (See S.1318 "The Elementary School Reading Emphasis Act of 1973," a bill sponsored by Senators Beall and Dominick). A recent survey of parents in the state of Maryland indicated that reading was the primary area in which the schools should concentrate.

Literacy Assessment

One area of particular concern is the assessment of

* Parts of the research presented herein were performed pursuant to research grants supported by the Catholic University of America and by Job Corps, Department of Labor, JCC No. 3294-99.

literacy. Traditionally literacy assessments were equated with measures of reading achievement. An individual was judged to be functionally literate if he received a grade level score on a standardized reading achievement test of anywhere from fifth to eighth grade level. Authorities differ as to which of these grade levels should be equated with functional literacy; but whatever the score, the measurement of literacy involved a standardized reading achievement test (usually designed for the elementary school child) and the assignment of a grade level score with norms developed based on performance of elementary school children.

The content and format of such tests usually fit the following pattern: reading comprehension is measured with a number of relatively short paragraphs, usually gradedin difficulty, followed by three or four questions designed to tap such skills as determining the main idea, noting details, and the like. The content of the paragraphs represents a range of interests suitable for a predominantly elementary school age child; practical reading tasks are usually not included. The student reads the paragraphs and responds to the questions by selecting one correct answer from a set of four and marking the appropriate answer blank. By implication, then, functional literacy, if measured by one of the standardized tests described

above, would be defined as the grade level score received on a test composed of such content.

It is suggested that tests such as these are not suitable measures of functional literacy, especially with adults and minority group members. The validity question is here considered in three aspects: the content of the tests; the format of administration; and the use and interpretation of scores.

Test Content

An examination of the content focuses, firstly, on the relationship between such content and content which might be considered more suitable or representative of the domain of functional literacy. Harris (1971) suggested that the focus of functional reading ability should be on reading skills required to cope with everyday experiences. Harris's two surveys (1970 and 1971) were built on this philosophy and contained practical reading content. In the development of the National Assessment material some emphasis was placed on including content which represented more practical aspects of reading, (National Assessment of Educational Progress, May 1972). Interest in practical reading material has likewise been expressed by such central figures in reading as Ruth

Love Holloway, Director of the Right to Read Program at the Office of Education, and Nathaniel Dixon, Executive Director of the National Reading Center. Although a firm definition of functional literacy has yet to be agreed upon, all indications are that emphasis will be on the performance of reading tasks directly related to practical real life experiences. It is thus suggested here that tests whose content is not representative of such practical real-life reading are inappropriate for the measurement of functional literacy.

Further, it is suggested that the content of many of the reading achievement tests is unsuitable for use with youth and adults. Such content is often child oriented and young adults are often poorly motivated to respond to what they consider to be a test "beneath" them. A somewhat related problem concerning the appropriateness of content is its suitability for minority group members. Surveys have indicated that large numbers of the illiterate in the United States are members of minority groups. Recognizing inappropriateness of content, Harcourt Brace has indicated that they are reviewing their tests with the view towards "soliciting their reactions /Blacks, Spanish-Speaking Americans, and others who are familiar with the needs and styles of pupils from a variety of

minority backgrounds/ to any content which might be, unintentionally, inappropriate or offensive for such children," (emphasis added) (Fitzgibbon, 1973, p. 3). The American Psychological Association, in their third draft of the Standards for Development and Use of Educational and Psychological Tests, cites social ills attributed to tests as one of the prime motivators in the revision of their Standards. They express concern with such areas as "failure to choose an appropriate test," (p. II).

Closely related to the type of information presented in the test is the physical appearance of such information. Thus, not only should the content be drawn from practical reading experiences, it should actually be as close in appearance to the actual material as possible. It is unclear what influence the physical layout of the material may have on a reader, but it is suggested that a representation which approximates the form in which the reader is likely to encounter the material in his actual reading provides a more accurate picture of the reader's true performance. For example, if one wanted to measure an individual's ability to read and interpret a lease, a facsimile of an actual lease rather than the content of the lease in some modified or rewritten version

should be presented. It would seem that this approach represents a more accurate approximation of the actual task. Face validity may have considerable effect on the student's attitude towards taking the test and the teacher's attitude towards interpreting its meaningfulness in the assessment of functional literacy.

Format of Administration

Test administration format may also affect student motivation to respond. Traditional reading tests may actually pose a threat for the poor reader or for the adult who has frequently faced failure in test situations. It remains unclear what effect the authority figure may have on student response, but it is suggested that such authority figure might create a feeling of anxiety which could adversely affect the student's ability to perform. Poor readers do not like to reveal their deficiencies, especially in group situations.

Most reading achievement tests are power tests, i.e., speed is not generally a factor influencing student performance. However, almost all reading tests are timed and the amount of time allotted for test completion may be insufficient for the poor reader. Thus an undue bias may result from administering a test in a group situation and imposing a time limit.

The test taking situation and the appearance of the test may have a negative effect on student performance. It has been suggested that students are "turned off" to taking tests, that such negative feelings may cause them to perform less well than they might in the actual situation. Thus, many tests, which are actually supposed to be samples of student performance, may not actually represent unbiased samples of tasks. Responses to such negatively viewed tests may not adequately reflect the student's true performance.

Response mode may also present bias in the testing of poor adult readers. If one is desirous of knowing a student's ability to perform a given reading task, then his actual performance should be measured--his output--rather than his ability to select one item from a set of four or five. The multiple choice format most frequently used in the testing of reading may actually prove to be an unreliable measure of an adult's literacy capability. The National Assessment materials and procedures support this approach as do some materials developed by HumRRO in their tests to assess job related skills, (National Assessment, 1972, and Sticht, 1972).

Use and Interpretation of Scores

A third area of concern in the assessment of functional literacy involves the use and interpretation of test scores. For the most part measurement of literacy has utilized norm referenced procedures whereby an individual's performance on a test is reported in terms of his relationship to others who have taken the test. One primary aim of norm referenced tests is to determine maximum discriminability among individuals and items are included which will maximize this discrimination. A grade level score refers to the degree to which an individual performs relative to others; it does not indicate the degree to which an individual performs relative to a standard or criterion of mastery of a given task. In literacy assessment, the question to be asked is whether or not an individual has mastered a sufficient amount of reading tasks which are representative of functional literacy. Reference is made to the amount and type of material mastered (i.e., whether or not the individual has achieved criterion) rather than how well he responds compared to others. The important consideration is whether or not the individual has demonstrated mastery of the stated tasks.

Other difficulties also occur with the use of grade

level scores in reporting performance of adolescent and adult groups. Many reading achievement tests were normed on elementary school children. Little meaning can be attached to the statement that an adult receives a grade level score of 4.2. This could mean either that he reads two months better than the average fourth grader at entrance into school or that he reads fourth grade material successfully but does not read fifth grade material as well. In any event he is being compared with fourth grade children on inappropriate material and thus the score's meaning is questionable.

Purpose of the Study

This study describes the research procedures used in the development and validation of R/EAL© (Reading/Everyday Activities in Life), a new test developed by the author to overcome some of the specific problems discussed above in assessing functional literacy among adolescents and adults. Specific objectives of this study are to:

1. Provide information about the design and development of R/EAL, including determination of reading criteria, establishment of task analyses, production and selection of individual test items, development of test format and procedures, and completion of final version of R/EAL.

2. Provide information about the validation of R/EAL, including procedures and data on reliability and validity for a select sample.

Design and Development of R/EAL

The following organizational scheme was followed in the construction of R/EAL: initial preparation and construction of the instrument, preliminary field testing, item analysis and selection, and production into its present form.

During the initial preparation and construction of the instrument, the author was guided by a number of conditions which attempted to overcome inadequacies of the tests. Firstly, the content must be representative of activities which could be considered directly related to practical life reading experiences. In addition, the content must be suitable for adolescents and adults. Further, the content must be presented in such a fashion as to closely resemble the appearance of the material as it usually is found. Since empirical information documenting frequently read material was not available at the time of the initial test development, the author selected materials and activities based on a logical and common sense approach. The identified areas included sets of directions, applications, technical documents, etc.

Secondly, it was decided to provide a test format that

would be motivating, eliminate administrator bias, have no time limits, and eliminate group test taking situations. Furthermore, many situations in which adolescents and adults are measured for functional literacy may not have trained test administrators, or may not have regularly scheduled classes which makes the usual group test administration procedures difficult. Also, the control of the test taking situation directly in the hands of the individual being tested appeared to be particularly desirable for adolescents and adults. As a response to these conditions, it was decided to use an individually controlled audio cassette input for all test directions, information and questions, and a booklet presenting facsimiles of various reading materials. In a study of self concept comparing audio-tape administration versus teacher administration, Giguere and Baker (1971) " ... found that a replicable method of test administration not subject to teacher mood or preparation will provide more representative data than one that varies from test occasion to test occasion," (pp. 9-10).

Thirdly, mode of student response needed to be determined. Questions which utilized a multiple response mode, while easier and less time consuming to score, do not allow the individual to demonstrate his ability to actually

perform certain tasks associated with the content. They may, in fact, only reveal the individual's ability to make intelligent choices from among a set of alternatives. The decision was thus made to use the student constructed, or open-ended, response mode on the assumption that it more accurately represented a student's ability to perform certain tasks.

Finally, the important decision was made to construct the test following procedures recommended by Nitko (1971). He suggests, in referring to the development of criterion-referenced tests that classes of behavior be defined, a set of test situations be specified, a representative sample of tasks be selected, and that the obtained score be capable of expressing the individual's performance characteristics in the classes. Task Analyses detailing the terminal objective and each of the enabling objectives necessary in the reading of a reading material were developed. Figure 1 is an example of one such analysis.

Insert Figure 1 About Here

Norm-referenced tests were considered unsuitable since they would not provide an indication of the student's ability to master functional literacy tasks, but rather his

ability in relation to others. Criterion-referenced, or content-referenced, tests provide information about how well the student has mastered the content of the test. The score interpretation was to be made directly to mastery of a predetermined cutoff point relating to a set of objectives. Glaser and Nitko suggest that "a criterion-referenced test is one that is deliberately constructed to yield scores that are directly interpretable in terms of specified performance standards," (in Nitko, 1971, p. 3).

Preliminary field testing was conducted during the academic year 1971-1972 on approximately three hundred individuals located in the Washington, D. C. area. The majority of the group were inner city disadvantaged high school students, although some seventh and eighth grade suburban students, and some students attending adult basic education classes were also included. The purposes of this field testing were (1) to determine if the content and format proved interesting, practical, and workable and what changes needed to be made; (2) to determine which questions needed to be eliminated or modified because of ambiquities, lack of clarity, and the like; (3) to determine which reading content and corresponding questions should be used in the revision of R/EAL.

For the most part, the procedures followed during the field testing were the same for all groups, although a few modifications had to be made depending on prevailing conditions in schools. Potential subjects were identified, arrangements made with appropriate officials for testing, and actual testing conducted. Since the test was administered via individually controlled tape recorder, the assembly of equipment presented additional problems, but these were overcome by the purchase of sufficient equipment. In most instances students were tested in small group situations with a member of the project staff present. Each student operated his own recorder and paced the input according to his needs. In a few cases one recorder was used for a small group and the input was paced by demands from the various group members to "repeat" or "stop." This procedure was abandoned early because it appeared to result in additional problems. In a few cases students were tested on an individual basis by a member of the project staff.

Since students were tested in a variety of situations it was not possible to obtain information about their reading ability, except gross judgment by teachers or counselors. For example, if a student was enrolled in an adult education program it was assumed that he had difficulty reading. In most

cases standardized test scores were unavailable and it was decided to forego a previously anticipated plan to explore the relationship between R/EAL test performance and performance on another reading test. Such steps were undertaken, however, during the validation phase.

The data collected during this field testing were subjected to a variety of statistical procedures, including the computation of p values or the proportion of students passing each item; means and standard deviations; and factor analyses. Since R/EAL was designed as a criterion-referenced test, the usual procedures for selection of items could not be followed. That approach would result in the use of items which would provide maximum discriminability of students rather than the use of items which would tap the predetermined content. For example, it might be desirable to include an item which measured one of the basic predetermined objectives yet that item might have a p value that did not provide maximum discriminability. In discussing the development of objective-based tests Giguere and Baker (1971) indicate a desire for but lack of procedural guidelines for the interpretation of data, (p. 10). Others assume that if an item measures an objective that is sufficient reason for its inclusion in a test.

Since precise guidelines for the selection of items for a criterion-referenced test are vague, the author decided to combine both a logical and empirical approach to determine the desired items.

Based on the information gleaned from this field testing phase, a revised version of R/EAL was constructed. Revisions included (a) lengthening the number of reading criteria presented in one booklet from seven to nine thereby lengthening the test from a thirty-five item to a forty-five item test, a procedure designed to tap both additional content areas and to increase reliability; (b) eliminating or modifying some questions which appeared to have unusual response patterns (e.g., the lower half of the distribution did better than the upper half) or which did not directly relate to the predetermined objectives; (c) shortening the initial directions presented on the audio cassette since informal observations revealed that such directions did not add to students' understanding of the tasks to be done; (d) altering the reading criterion in a few cases by decreasing the length of the material presented, eliminating ambiguous portions, or adding additional information such as a title or heading; (e) producing the audio input by a trained professional in a sound

studio to eliminate extraneous sound or difficulties in comprehending the spoken word.

Thus the present revision of R/EAL (a) contains reading criteria selected from such common daily reading activities as food store advertisements, directions for preparation of food, want ads, leases, maps, etc.; (b) includes tasks related to each criterion developed based on the reading functions required to deal with the individual criteria. Such tasks have been translated into specific questions following guidelines recommended by Davis (1971), "item analysis data ... /were/ used as a basis for refining the items through insightful editing, but the use of item-test correlation coefficients or difficulty indexes ... /were/ not allowed to affect the validity of the test by distorting the proper representation of behavior categories ... " (p.1); (c) presents all directions and questions in aural form via individually operated cassette players.

Validation of R/EAL

This section deals with the methodology used for validating R/EAL, including sample selection, testing procedures, and item analysis. Data relating to individual item and total test statistics, factor analysis data, reliability,

and validity are reported.

Methodology

Subjects. The subjects selected for the validation sample were students enrolled in a residential manpower training program servicing disadvantaged youth ages 16 to 21. Most of the enrollees were high school dropouts, many of whom lack basic skills in reading. For the most part, Job Corps enrollees are Black, Mexican-American, rural white, Puerto Rican or American Indian. To be eligible for the program their families must be at the poverty level.

Four Job Corps Centers, two male and two female, were selected for participation in the program. (Job Corps has only a very few Centers which are coeducational.) The Centers were selected to represent the various groups served in the Job Corps program. Their geographic locations included sites in New Jersey, West Virginia, New Mexico and Texas. Total numbers of enrollees varied from Center to Center.

Selection of subjects at each Center varied depending on conditions existing at the Centers during time of testing. In two Centers a random sample (one computer-generated) was identified. An attempt was made to select a random sample in a third Center but this was not entirely possible. Because of

conditions at the fourth Center whole classes of enrollees were used for testing. Although procedures for sample selection varied, it did not appear that any known bias was introduced in the selection of subjects. Since students in two Centers were of Mexican-American descent, a restriction was placed that only those fluent in English could participate. (Another study was conducted with information transmitted in Spanish, but is not the subject of this paper.) Table 1 reports the sample size for each Center, the total size, and the distribution by sex.

Insert Table 1 About Here

Testing Procedures. At all Centers testing was supervised by the author, with the assistance of one or two others on her staff. Centers also provided personnel who were available to assist with equipment, scheduling, and the like.

All enrollees were tested in a special room designated for testing. About twenty-five students were tested at a time. Each student had his own individually controlled and operated cassette recorder and earphone and was permitted to work at his own pace. Upon completion of the test he returned to his normal pursuits and a new enrollee took his place. Thus testing

proceeded almost on continuous basis (except for a lunch break). In some Centers over a hundred enrollees were tested in one day.

Directions for all students were the same and consisted only of explaining the use of the tape recorder and earphone; prior to the actual test date, however, students had been told that they would participate in a new type of literacy testing.

Student responses were scored correct or incorrect according to a predetermined objective scoring key. Partial or ambiguous responses were judged incorrect. All scoring was carried out by the author or members of her staff.

Data Analysis. All pertinent enrollee data were transferred to IBM cards for use in the data analysis. All statistical analyses were computed on an IBM 360 computer.

Magnuson (1967), in discussing the usefulness of item analysis procedures, indicated the relationship of item analysis techniques to the questions of reliability and validity. He suggested that the dependability of an obtained score is "an estimate of his true score (i.e., the reliability of data), which determines the value of the test. The reliability and validity of the data depend on the properties of the individual items which make up the test," (p. 197). Item analysis statis-

tics were computed on available data. These included p values (proportion of subjects passing an item) for total group, upper half, and lower half; inter-item correlations, point biserial correlations, and factor analyses.

Results

This section provides specific information obtained from the above-described sample.

Item Analysis. The information provided in Table 2

Insert Table 2 About Here

---

indicates the difficulty level of each item in the test for the sample described above. A high p score reflects an easy item; a low p score the reverse. If we are concerned with discriminability of items, a high p score would be a poor indicator of discriminability. But this test is concerned with mastery to a predetermined set of objectives. Thus, a high p score would suggest that a large number of the group tested had mastered that item.

Additional information about difficulty levels of individual items can be obtained by examining the proportion of the upper half of the distribution that passed a given item and the proportion of the lower half that passed the same item. If the item is sound, it would be anticipated that a higher proportion of the upper half than the lower half would pass the item. In addition to examining the item difficulty,

item analysis procedures also call for correlations of items with total test scores. Guilford (1965) suggests that item-test correlations are more important than difficulty of individual items because they indicate whether or not a test item discriminates in line with other items in the test. For this type of analysis Guilford suggests the use of a point biserial correlation. Magnuson (1967) suggests that the magnitude of the point biserial correlation is greatly affected by the difficulty level of an item. This results in "very easy or very difficult test items (having) systematically lower coefficients for the correlation with the test than items of medium difficulty," (p. 209). Recognizing that the difficulty level of a test item affects the correlation, caution in its interpretation must be exercised. Again, it must be emphasized that certain items which were logically included based on their relationship to the predetermined tasks and objectives may be at the extreme difficulty levels. Table 3 reports the point biserial correlations between individual test items and total score.

Insert Table 3 About Here

Additional statistics about the items were ...
... particular the correlation of items with other items ...
test. Table 4 reports these correlations for the forty-s...
items used in R/EAL. In interpreting these intercorrelat...

Insert Table 4 About Here

---

it should be noted that a number of aspects of reading are being measured and that reading is not viewed as a unitary trait.

Needing additional information about the relationship of each item to the others necessitated a factor analysis to be performed. A principal components analysis with a varimax rotation program was utilized. These calculations yielded three factors which accounted for 100% of the common variance but only 32% of the total variance. Table 5 gives the information obtained in the analysis.

Insert Table 5 About Here

Reliability. Estimates of reliability were calculated for R/EAL and are reported herein. A number of procedures have been developed for estimating the reliability of a test. In particular, special procedures for estimating the reliability of a content-referenced or mastery test have been proposed. These procedures are still in experimental stages,

however. Livingston (1972) offered a reliability coefficient based on deviations of scores from the criterion score rather than deviations from the mean. Harris (1972), in response to Livingston's work suggested that " ... his work fails to advance reliability theory for the special case of criterion-referenced (content-referenced) testing," (p.29). Marshall (1973), too, suggested difficulties with Livingston's coefficient. Rather, he offered additional information related to the methodology of determining reliability of criterion-referenced tests. He suggested three indices to be used in the estimate of reliability: index of efficiency, index of sensitivity of instruction and index of separation. Since Marshall's work is still highly experimental, however, it was decided not to pursue these coefficients at this time.

Fluidity of thought concerning acceptable procedures for estimating reliability of mastery tests had led the author to select classical internal consistency measures for estimating the reliability of R/EAL. Kuder-Richardson 20 (KR20) procedures were employed to provide measures of "both equivalence (of items) and homogeneity," (Anastasi, 1961, p. 122). Table 6 offers the results of the calculations.

Insert Table 6 About Here

Validity

Questions of validity need to be considered in determining if a test is appropriate for a particular use. Validity refers to the degree to which a test actually measures what it purports to measure, (Anastasi, 1961, p. 29). Further, validity coefficients provide a check on how well the test fulfills its function.

The APA Standards (1973) also considers the question of validity of a test. In their terms validity is concerned with the accuracy of the information that can be inferred from the test score. The measuring instrument "is an operational definition of a specified domain of skill or knowledge," (p. VI). Information related to the operational definition of functional literacy has been supplied earlier in the discussion of the rationale and development of R/EAL.

In the discussion here, two types of validity will be considered: criterion-related validity and content validity. Criterion-related validity refers to the relationship of this test to some other (external) criterion designed to measure the same function. Criterion-related validity was determined by selecting a standardized reading achievement test and computing a correlation between R/EAL and the

reading achievement test. For the target population described above, the Stanford Achievement Test was selected. The Pearson Product Moment correlation between the two tests was .74 (n=434) and the standard error of measurement was equal to 5.28.

Content validity is another suitable area when exploring the validity of R/EAL. A demonstration of content validity must show that the behaviors sampled in the test are a representative sample of behaviors from the universe of behaviors. Giguere and Baker (1971) suggest that the validity of criterion-referenced tests "does not depend on a series of highly related correlations but rather on the user's acceptance of the specified premises upon which the instruments are based," (p.2). Reference is made to the Task Analysis at page 29. Each question used in R/EAL is selected directly from the Task Analysis and is so designed as to represent as much of the domain of tasks as is possible. These Analyses specify objectives of the test and indicate how the "component tasks make up the total domain," (Standards, 1973, p. VII).

Implications

The need for a test concerned with the practical application of reading in daily life is great. R/EAL

attempts to overcome some of the problems inherent in tests which are presently in use. Reliability and validity figures on R/EAL tend to support its use as a viable assessment instrument for functional literacy.

Item analysis figures show a difficulty level of items for the total group ranging from .35 to .97 with a median of .60. Since items were designed to reflect predetermined objectives and since it was known from other information that the sample reflected a range of reading abilities, this variation in response to individual items would be anticipated. An examination of differences between the upper and lower half of the distribution reflects, as anticipated, that the upper half had mastered more of the tasks than the lower half.

The point biserial correlations suggest that some items are more closely related to the total score than others; but the restrictions suggested by Magnuson must be taken into account in interpreting these correlations. The intercorrelation matrix (Table 4) and the factor analysis (Table 5) lend further support to the hypothesis that functional literacy is not a unitary trait and may be influenced by such factors

as the content or format of the material. In addition, the factor analysis data seem to suggest that at least three separate factors are measured by R/EAL. Items in Factor I come primarily from tasks relating to reading road maps and road signs. Factor III items are all from the job application. The heaviest loading is on Factor II and represent all other items from the test. It is interesting to note that, in a related study, the Stanford Achievement Test split fairly evenly on Factors I and II.

Areas for future investigation are of special importance in a test of this nature. Research relating to pre-test, post-test differences is currently being completed and will be reported subsequently. Additional research is also underway using R/EAL with a population of deaf students. In that study the audio portion is being translated into a videotaped total communication presentation. Additional research with other populations and age groups is also being considered. The use of R/EAL as a diagnostic/prescriptive instrument was part of the original design but empirical validation of these procedures still needs to be undertaken. Finally, equivalent forms of R/EAL and the use of additional reading criteria are being developed.

ROAD SIGNS

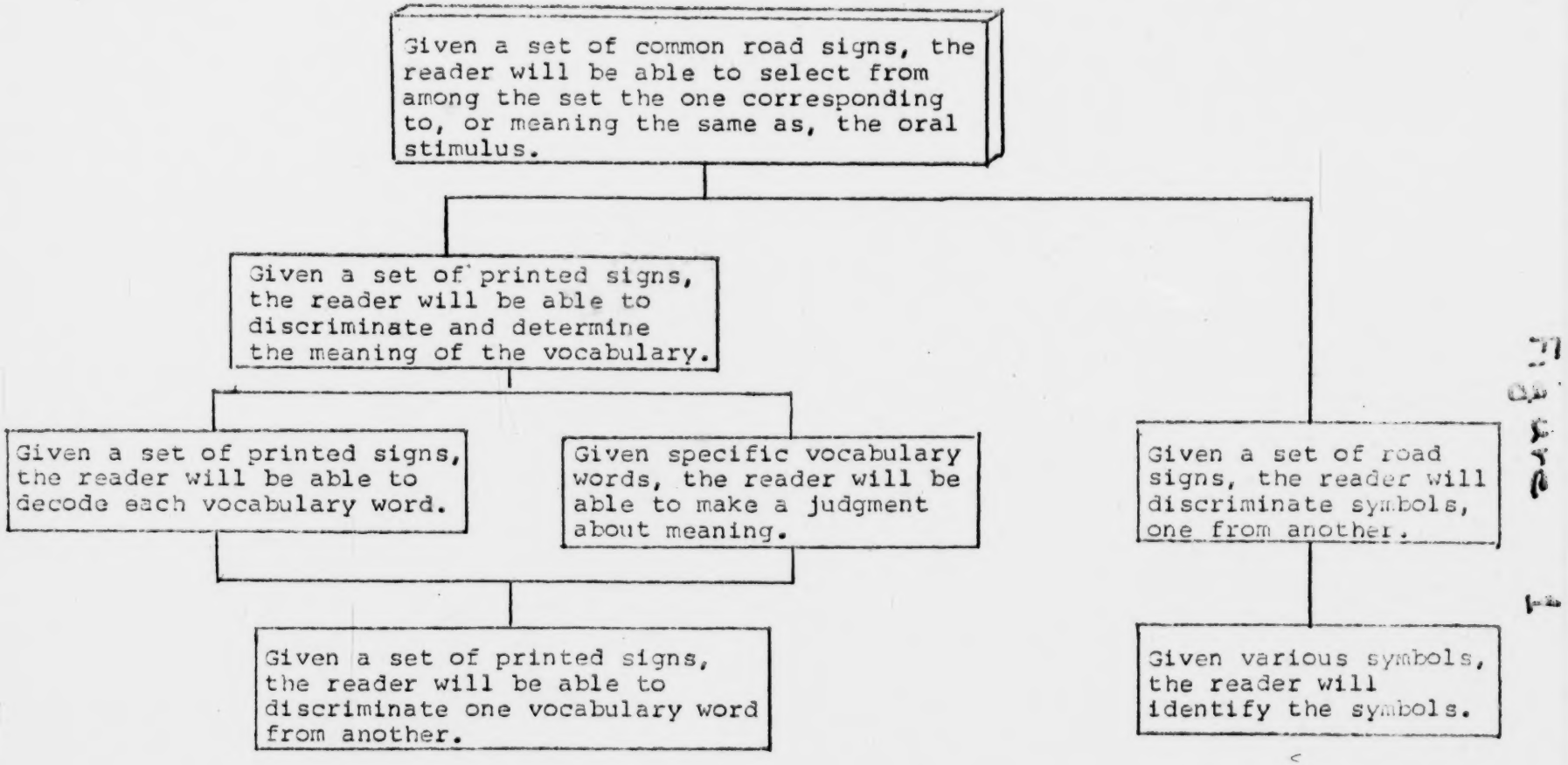
Given a set of common road signs, the reader will be able to select from among the set the one corresponding to, or meaning the same as, the oral stimulus.
Given a set of printed signs, the reader will be able to discriminate and determine the meaning of the vocabulary.
Given a set of road signs, the reader will discriminate symbols, one from another.
Given a set of printed signs, the reader will be able to decode each vocabulary word.
Given specific vocabulary words, the reader will be able to make a judgment about meaning.
Given a set of printed signs, the reader will be able to discriminate one vocabulary word from another.
Given various symbols, the reader will identify the symbols.

Table 1

Sample size by Center & Sex

Center	Size*	Sex	Size
1	98	Male	169
2	101	Female	265
3	164		
4	71		
Total	434		434

*To a general degree these sizes and sex distributions reflect a proportion of the total size and sex distributions at the Centers.

Table 2

Item Difficulty for Total Group, Upper and Lower Halves (n = 434)

Item No.	Pupper	Plower	Ptotal
1	.98	.78	.88
2	.84	.58	.71
3	.98	.77	.88
4	.98	.79	.89
5	.80	.37	.58
6	.99	.76	.87
7	.66	.46	.56
8	.55	.26	.41
9	.51	.21	.36
10	.53	.18	.35
11	.94	.62	.78
12	.93	.49	.71
13	.72	.29	.50
14	.68	.22	.45
15	.62	.14	.38
16	.86	.49	.68
17	.82	.45	.63
18	.81	.40	.60
19	.50	.17	.36
20	.71	.30	.51
21	.99	.95	.97
22	.86	.48	.67
23	.93	.67	.80
24	.86	.47	.66
25	.59	.27	.43
26	.91	.37	.64
27	.90	.43	.67
28	.80	.37	.58
29	.92	.53	.72
30	.82	.24	.53
31	.71	.28	.49
32	.82	.27	.54
33	.97	.53	.75
34	.71	.28	.50
35	.75	.26	.50
36	.73	.29	.51
37	.83	.38	.61
38	.84	.27	.56
39	.79	.28	.54
40	.76	.25	.51
41	.99	.94	.96
42	.98	.91	.95
43	.82	.46	.64
44	.94	.61	.77
45	.77	.19	.48

Table 3

Point-Biserial Correlations

Individual Test Items to

Total Test Score

Item Number	Correlation	Item Number	Correlation
1	.48	26	.66
2	.37	27	.60
3	.49	28	.51
4	.47	29	.63
5	.55	30	.64
6	.56	31	.47
7	.33	32	.57
8	.39	33	.65
9	.37	34	.48
10	.44	35	.55
11	.47	36	.53
12	.59	37	.59
13	.48	38	.66
14	.55	39	.62
15	.52	40	.62
16	.50	41	.25
17	.51	42	.25
18	.55	43	.45
19	.45	44	.53
20	.47	45	.63
21	.27		
22	.56		
23	.47		
24	.55		
25	.43		

Table

R/SAI Intercorrelation Matrix (n = 434)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
1		26	50	35	31	43	16	14	12	15	19	31	19	18	19	19	20	28	19	15	25	23	20	29	17	32	27	27	31	27	27	20	24	25	25	19	26	30	35	28	05	05	12	25	29	1
2			30	14	24	26	-01	15	12	12	05	10	14	16	19	15	18	15	18	20	11	24	13	15	18	18	16	11	19	21	30	20	20	18	22	20	18	21	23	25	-01	01	08	17	22	2
3				42	37	39	19	13	12	13	20	26	20	21	28	22	29	24	22	17	15	26	23	28	19	26	24	24	28	23	30	26	36	25	25	21	20	25	28	27	08	07	11	29	28	3
4					30	32	09	12	17	20	16	28	14	26	22	30	21	23	14	19	12	28	21	29	19	31	31	25	33	28	19	18	30	18	21	23	23	31	27	29	05	05	17	24	27	4
5						30	16	18	10	26	13	25	19	26	23	31	27	20	26	22	14	27	21	23	17	34	32	31	34	28	40	31	36	32	32	23	25	32	28	30	12	15	27	21	33	5
6							26	25	18	20	31	33	23	25	19	24	33	23	15	23	32	28	21	26	25	37	39	30	37	29	26	28	42	21	30	28	30	36	34	34	04	07	18	27	28	6
7								14	11	23	11	23	13	17	13	09	25	08	06	11	10	14	06	14	17	20	20	13	22	09	08	12	21	18	21	09	21	25	14	12	06	03	14	18	22	7
8									31	20	18	20	22	27	19	16	16	19	08	19	14	19	21	18	22	24	24	09	19	22	13	26	21	10	16	15	15	21	20	19	03	06	20	21	23	8
9										17	12	23	16	27	16	23	14	16	09	14	07	18	17	15	19	22	18	14	20	23	09	21	20	12	19	24	23	18	19	15	04	04	20	15	20	9
10											16	24	22	17	22	22	20	23	24	14	07	19	15	20	23	29	31	21	26	27	21	31	27	19	23	16	23	20	18	24	03	04	16	16	26	10
11												46	36	30	26	13	19	27	20	21	18	28	25	31	21	30	30	23	30	22	10	26	37	13	17	25	30	32	28	24	08	05	23	24	26	11
12													34	37	34	24	27	34	24	23	14	33	33	39	15	41	35	28	45	38	18	33	39	21	25	29	36	38	31	28	13	11	25	29	31	12
13														32	31	23	24	25	20	17	09	23	25	34	20	28	22	18	21	25	18	29	28	18	26	16	24	29	24	28	09	07	16	24	27	13
14															31	22	24	25	28	24	04	27	24	28	27	37	29	26	37	36	14	33	34	20	28	27	30	37	31	33	07	08	24	28	32	14
15																19	25	28	28	21	07	25	26	27	15	32	23	21	26	30	26	33	34	27	32	28	28	29	28	33	10	07	20	22	34	15
16																	23	24	15	26	19	23	20	31	21	42	31	32	21	25	21	26	25	23	34	27	30	29	27	21	17	22	18	25	29	16
17																		37	23	24	13	23	16	24	22	26	32	19	31	30	20	27	31	18	26	24	27	36	31	30	14	13	20	27	35	17
18																			28	28	15	30	24	29	20	32	31	29	33	38	24	28	36	22	23	37	27	34	32	32	16	13	19	30	32	18
19																				16	07	24	20	22	19	26	18	20	25	35	21	24	25	25	16	26	22	29	29	37	09	11	17	19	28	19
20																					14	27	26	20	09	35	26	24	30	33	12	19	28	14	23	31	29	25	22	26	14	15	23	33	36	20
21																						21	20	15	09	14	15	14	21	07	14	10	19	08	11	11	15	13	13	17	20	22	11	18	08	21
22																							36	41	29	29	30	22	38	39	23	26	44	26	28	36	32	29	36	31	14	11	18	24	29	22
23																								32	19	26	30	24	38	31	16	19	30	10	25	24	29	21	26	26	03	07	12	29	26	23
24																									29	42	37	25	34	33	17	33	33	22	27	26	30	33	36	26	13	15	19	25	24	24
25																										21	21	17	20	25	13	26	30	10	21	19	31	29	23	24	06	07	18	19	24	25
26																											59	45	44	40	24	40	43	37	34	29	33	44	29	38	17	18	25	29	37	26
27																												44	46	42	18	29	38	28	33	30	30	33	32	33	08	06	22	28	31	27
28																													33	35	24	23	31	22	28	19	27	26	25	30	15	13	21	22	30	28
29																														42	20	32	47	23	26	33	38	39	37	37	17	16	22	36	35	29
30																															28	33	38	27	30	40	36	37	42	42	18	16	28	35	42	30
31																																40	28	39	38	22	28	27	33	29	06	06	17	16	19	31
32																																	42	34	30	23	40	40	28	30	16	12	25	23	28	32
33																																		39	34	26	35	41	36	37	15	16	24	33	31	33
34																																			31	15	24	32	34	31	11	08	14	15	31	34
35																																				32	23	33	32	34	09	11	24	24	37	35
36																																					33	37	40	40	09	11	18	27	32	36
37																																						45	42	37	13	11	26	30	37	37
38																																							57	50	16	15	26	30	49	38
39																																								51	15	14	22	30	42	39
40																																									12	09	26	27	47	40
41																																										82	25	35	18	41
42																																											26	30	20	42
43																																												39	44	43
44																																													44	44
45																																														45

Table 5

Factor Loadings for Items in R/EAL*

Item Number	Factor I	Factor II	Factor III
1	.49		
2	.42		
3	.52		
4			
5	.52		
6	.44		
7			
8			
9			
10			
11		.52	
12		.60	
13		.43	
14		.57	
15		.41	
16			
17			
18		.44	
19			
20			
21			
22		.44	
23		.46	
24		.48	
25			
26		.53	
27		.52	
28			
29		.58	
30		.54	
31	.64		
32		.40	
33		.51	
34	.48		
35	.42		
36		.42	
37		.49	
38		.51	
39	.42	.43	
40		.47	
41			.73
42			.72
43			
44		.50	.43

*Only those loadings which were .40 or over are reported.

Table 6

Reliability of R/EAL in

Target Population (KR-20)

$n = 434$ $\bar{X} = 28.09$ $s = 10.36$ $r_{11} = .93$

$s_e = 2.75$

References

Allen, James E. The Right to Read: Education's New National Priority. The American School Board Journal, 157, October 1969, 25-27. Also as The Right to Read--Target for the 70's. Journal of Reading, 13, No. 2, Nov., 1969, 95-101.

American Psychological Association. Standards for Development and Use of Educational and Psychological Tests. third draft, 1973.

Anastasi, Anne. Psychological Testing. New York: The Macmillan Co., 1969.

Beall, J. Glenn and Dominick, Peter. The Elementary School Reading Emphasis Act of 1973. S. 1318. Congressional Record, Vol. 119, No. 45, March 22, 1973.

Davis, Fred B. Criterion-Referenced Measurement. ERIC Clearinghouse on Tests, Measurements, and Evaluation, 1971 AERA Summaries, March 1972.

Edmonston, Leon and Randall, Robert. A Model for Estimating the Reliability and Validity of Criterion-Referenced Measures. Paper presented at the Annual Meeting of AERA, 1972.

Fitzgibbon, Thomas J. The Use of Standardized Instruments with Urban and Minority-Group Pupils. Test Department, Harcourt, Brace Jovanovich, Inc., n.d.

Giguere, Constance L. and Baker, Eva L. The Development of Criterion-Referenced Tests of Self-Concept and Attitude Toward School. University of California, Los Angeles, 1971, ERIC Document ED 054 192.

Green, Donald Ross. Racial and Ethnic Bias in Test Construction. Monterey, California: McGraw Hill, adapted from Final Report of United States Office of Education, Contract No. OEC-9-70-0058(057), n.d.

Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill, 4th edition, 1965.

Harris, Chester. An Interpretation of Livingston's Reliability Coefficient for Criterion-Referenced Tests. Journal of Educational Measurement. Vol. 9, No. 1, Spring 1972, 27-29.

Harris, Louis and Associates, Inc. Survival Literacy Study. Conducted for the National Reading Council. September 1970, mimeo.

_____. The 1971 National Reading Difficulty Index: A Study of Functional Reading Ability in the United States. Conducted for the National Reading Center. August 1971, mimeo.

Lichtman, Marilyn. The Functional Literacy Study: Assessment and Interpretation of Literacy Levels and Reading Tasks of Selected Job Corpsmembers. A proposal submitted to the Department of Labor, July 1972, mimeo.

Livingston, Samuel A. A Classical Test-Theory Approach to Criterion-Referenced Tests. Paper presented at the Annual Meeting of AERA, 1972.

_____. Criterion-Referenced Applications of Classical Test Theory. Journal of Educational Measurement. Vol. 9, No. 1 Spring 1972, 13-25.

Magnuson, David. Test Theory. Reading, Mass.: Addison-Wesley, 1967.

Marshall, J. Laird. Reliability Indices for Criterion-Referenced Tests: A Study Based on Simulated Data. Paper presented before the National Council of Educational Measurement, Feb. 1973.

National Assessment of Educational Progress, Reading--Selected Exercises, Report 02-R-20, Educational Commission of the States, Suite 300, Denver Colorado 80203, May, 1972.

_____. Reading Summary, Preliminary Report 02-R-00, May 1972.

Nitko, Anthony J. A Model for Criterion-Referenced Tests Based on Use. Paper presented at Annual Meeting of AERA, Feb. 1971, ERIC Document ED 049 318.

Sharon, Amiel T. Reading Activities of American Adults. Princeton: Educational Testing Service, December 1972.

Sticht, Thomas, et al. Determination of Literacy Skill Requirements in Four Military occupational Specialties. Alexandria: Human Resources Research Organization, Technical Report 71-23, November 1971.

Sticht, Thomas G. and Caylor, John S. Development and Evaluation of Job Reading Task Tests. Journal of Reading Behavior, Vol. 4, No. 4, Fall 1972, 29-50.